CS6923 Machine Learning, Fall 2018
Prof. Hellerstein, NYU School of Engineering

Homework 3 - Solutions (14 points)

Part I: Written Exercises (6 points)

1. (a) (1 point) The first dataset with 4 positive examples and 5 negative examples has higher entropy.

   (b) (1 point)

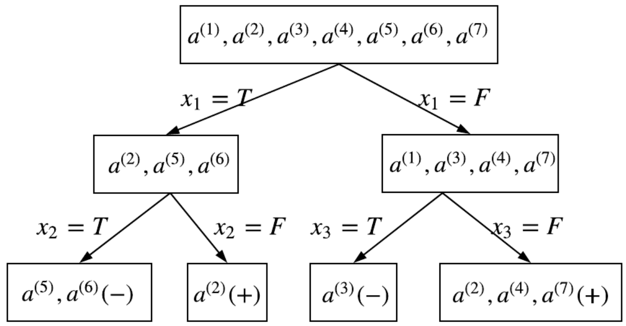
   $$\begin{aligned}&\text{Information-Gain}(x_1)\\ &= Entropy(x_1) - (p(x_1 = F)Entropy(x_1 = F) + p(x_1 = T)Entropy(x_1 = T))\\ &= -\frac{4}{7}log_2\frac{4}{7} - \frac{3}{7}log_2\frac{3}{7} - (\frac{4}{7}(-\frac{1}{2}log_2\frac{1}{2} - \frac{1}{2}log_2\frac{1}{2}) + \frac{3}{7}(-\frac{1}{3}log_2\frac{1}{3} - \frac{2}{3}log_2\frac{2}{3}))\\ &= 0.02\end{aligned}$$

   (c) (2 points) There may exist ties for entropies, so this question has more than one answers. Here is one sample solution.

   Note: You will lose one point for this question if you only draw the decision tree without calculating the entropies/information gains.

   Dataset: $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}, x^{(6)}, x^{(7)}$:
   Information Gain of choosing $x_1 : 0.128$
   Information Gain of choosing $x_2 : 0.128$
   Information Gain of choosing $x_3 : 0.128$
   Choose x_1 and divide the examples into $S_1 : x^{(1)}, x^{(3)}, x^{(4)}, x^{(7)},\ S_2 : x^{(2)}, x^{(5)}, x^{(6)}$
   Dataset: $x^{(1)}, x^{(3)}, x^{(4)}, x^{(7)}$:
   Information Gain of choosing $x_2 : 0.123$
   Information Gain of choosing $x_3 : 0.811$
   Choose x_3 and divide S_1 into $S_3 : x^{(1)}, x^{(4)}, x^{(7)},\ S_4 : x^{(3)}$
   Dataset: $x^{(2)}, x^{(5)}, x^{(6)}$
   Information Gain of choosing $x_2 : 0.918$
   Information Gain of choosing $x_3 : 0.252$
   Choose x_2 and divide S_2 into $S_5 : x^{(5)}, x^{(6)},\ S_6 : x^{(2)}$

   The final decision tree is:



(d) (1 point)

$$H(Y) = -\frac{4}{7}log_2\frac{4}{7} - \frac{3}{7}log_2\frac{3}{7} = 0.985$$

$$\begin{aligned} H(Y|X) =& P[X=F]*(-P(Y=+|X=F)log_2P(Y=+|X=F) \\ & - P(Y=-|X=F)log_2P(Y=-|X=F)) \\ & + P[X=T]*(-P(Y=+|X=T)log_2P(Y=+|X=T) \\ & - P(Y=-|X=T)log_2P(Y=-|X=T)) \\ =& 0.965 \end{aligned}$$

$$H(Y) - H(Y|X) = 0.02$$

(e) (1 point) $z \cdot (-\frac{1}{z}log_2\frac{1}{z}) = -log_2\frac{1}{z}$

Part II: Programming and Questions (8 points)

2. (2 points)

(a) (0.5 point) $k^2 - 9k - 84$

(b) (0.5 point) $k_1 = -5.71, k_2 = 14.7$

(c) (0.5 point)
for $k_1 = -5.71$, $[-0.926, 0.378]$
for $k_2 = 14.7$, $[0.691, 0.72]$

(d) (0.5 point)
eigenvalues = [-5.71028893, 14.71028893]
eigenvectors = [[-0.9259401 , -0.6894021], [0.37767039, -0.72437887]]

3. (2 points)

(a) (0.5 point) $B = \begin{bmatrix} -0.75 & -1.5 & 0.5 \\ 3.25 & 2.5 & 0.5 \\ 1.25 & -2.5 & -3.5 \\ -3.75 & 1.5 & 2.5 \end{bmatrix}$

(b) (0.5 point) $s_{1,3} = -4.167$

(c) (0.5 point) 12.9788

(d) (0.5 point) $\begin{bmatrix} 0.26018674 & -1.41900435 \\ -0.87353472 & 4.03721245 \\ -4.04749635 & -1.8486773 \\ 4.66084433 & -0.7695308 \end{bmatrix}$

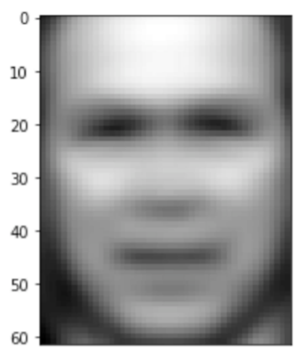
5. (4 points)

(a) (0.5 point)

plt_face(fea[3])

(b) (0.5 point)

plt_face(np.mean(fea, axis=0))



(c) (1 point)

import sklearn.decomposition as skd

pca = skd.PCA(n_components = 5)

skd.PCA.fit(pca, fea)

W1 = pca.components_

W = W1.transpose()

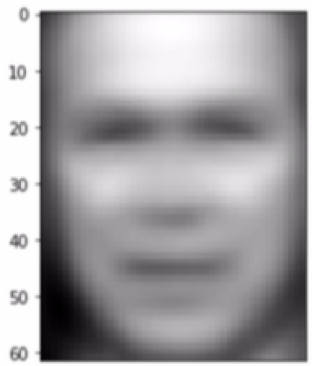
Z = pca.transform(fea)

Z[3]

Result: [202.5423 -261.47693 418.9748 -29.398222 39.78391]

(d) (2 points)

```
pca = skd.PCA(n_components=5)
skd.PCA.fit(pca, fea)
W1 = pca.components_
W = W1.transpose()
Z = pca.transform(fea)
plt_face(np.dot(W, Z[3]) + np.mean(fea, axis=0))
```




```
pca = skd.PCA(n_components=50)
skd.PCA.fit(pca, fea)
W1 = pca.components_
W = W1.transpose()
Z = pca.transform(fea)
plt_face(np.dot(W, Z[3]) + np.mean(fea, axis=0))
```